w o r k i n g

p a p e r

12 25

Learning and Occupational Sorting

Jonathan James

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 12-25 November 2012

Learning and Occupational Sorting

Jonathan James

This paper develops and estimates a model of occupational choice and learning that allows for correlated learning across occupation specific abilities. In the labor market, workers learn about their potential outcomes in all occupations, not just their current occupation. Based on what they learn, workers engage in directed search across occupations. The estimates indicate that sorting occurs in multiple dimensions. Workers discovering a low ability in their current occupation are signicantly more likely to move to a new occupation. At the same time, workers discovering a high ability in some occupations are more likely to move up the occupational ladder into managerial occupations. By age 28 this sorting process leads to an aggregate increase in wages similar to what would occur if all workers were endowed with an additional year of education.

Keywords: Occupational Choice, Correlated Learning,Wage Dynamics.

JEL Codes: D83, J24, J62.

Jonathan James is at the Federal Reserve Bank of Cleveland and can be contacted at jwj8@duke.edu. Updated versions of this paper are also available at: http://econ.duke.edu/~jwj8. The author is grateful to Peter Arcidiacono and Fabian Lange for helpful comments.

1 Introduction

Models of uncertainty and learning have provided an important framework for studying the rapid occupational mobility and early career wage growth of younger workers. Two notable approaches have been developed in this setting. The first, as illustrated in Miller (1984), is the occupational matching model whose main emphasis is experimentation. In this model, workers do not know how productive they are in different occupations. They spend the early years of their career searching for their comparative advantage by sorting across occupations. Important to this approach is that occupational productivities are unrelated, so poorly matched workers can always do better by changing occupations, while highly matched workers never change occupations.[1]

A second approach, exemplified in Jovanovic and Nyarko (1997), forms a different view of learning, which characterizes a stepping-stone model of occupational mobility. Under this approach, occupations can be ordered along an occupational ladder. Generally, occupational productivities are assumed to be perfectly correlated, so as learning occurs in entry level occupations, successful performance propels workers to occupations on higher rungs. This view emphasizes vertical mobility from low skill occupations to high skill occupations, where poorly matched workers are likely to stay in entry level jobs and highly productive workers are the most mobile.[2]

The goal of this paper is to develop and estimate a model of occupational choice that incorporates these multiple dimensions of learning into a single framework. The model nests both sorting as experimentation (as emphasized by Miller (1984)) and sorting through promotion (cf. Jovanovic and Nyarko (1997)) and the relative strength of these effects will be estimated. In this way, the worker's optimal search strategy becomes an endogenous

[1] Empirical work that assumes independence of occupation specific productivities has been conducted in McCall (1990), Neal (1999), and Pavan (2009). Independence of matches is also a typical feature in industrial organization models of experience goods, for example Ackerberg (2003) and Crawford and Shum (2005).

[2] Gibbons et al. (2005) use an IV approach to estimate a model of learning about absolute advantage that resembles this framework. Early work in Shaw (1987) explores these types of mobility patterns using proxies for the transferability of skills across occupations.

feature of the model rather than imposed by assumption.

In the model, individual productivity in different occupations is arbitrarily correlated, so information the worker receives about their productivity in one occupation is informative about their productivity in other occupations. Agents learn by choosing their occupations optimally. Thus learning will be multidimensional and the optimal search strategies are complex functions of the unknown parameters of the model and the worker's complete history of information. This more general framework produces a very large state space, which creates well known computational challenges for estimating dynamic discrete choice models. These issues are compounded by the high-dimensional nature of the occupational choice set, imposing significant computational costs.[3]

This paper overcomes these computational challenges by implementing a new method for estimating correlated learning models. In addition to being able to tractably estimate these complicated models, this approach has two other advantages. First, it can be carried out with mild assumptions on the learning process. While it is assumed that workers are Bayesian learners, the formation of beliefs is incorporated in a flexible way that does not necessarily require the econometrician to know the exact probability density function in which beliefs are based on. The second benefit of the estimator is that it easily accommodates a large choice set and state space, allowing for a finer level of aggregation of occupations in estimation.

The model is estimated on a subsample of the 1997 cohort of the National Longitudinal Survey of the Youth (NLSY97). Two main conclusions are drawn from this analysis. First, for most occupations, by age 28 the conditional mean of ability for those actually employed in that occupation is higher than the unconditional mean. For example, the average Construction ability for those choosing Construction at age 28 is 6% higher than if workers chose occupations randomly, all else equal. In many cases this sorting is significant relative to other factors affecting wages. For example, the

[3] Papgeorgiou (2010) develops a model where abilities are transferable across occupations but only allows two occupations in the empirical model citing computational constraints.

change in wages associated with workers sorting into occupations where they have higher ability are comparable to the total increase in wages that would occur if all workers were endowed with an additional year of education.

The second conclusion is that the empirical hazard of choosing an occupation does not always increase monotonically in ability. In many occupations the most productive workers are the most likely to leave these occupations and move to Manager occupations. For example, entry level Construction workers who transition to Manger occupations by age 28 are on average 21% more productive Construction workers and 17% more productive Managers compared to their cohort. On the extreme, some occupations are characterized by negative sorting, where high ability workers are the most likely to change occupations and low ability workers are the most likely to go to these occupations, causing the average ability in these occupations to decrease over time. Food Service and Office and Administrative Support occupations fall into this category for the population studied.

An important implication of the multidimensional selection underlying the data is the direction of the bias that occurs when this selection is not accounted for. The standard interpretation of ability bias is that OLS estimates will overstate the returns to occupational tenure. There is clear evidence of this from a comparison of the estimated returns to occupation tenure from the model with OLS estimates. For Sales occupations, the OLS estimates of the return to Sales experience is 50% higher than the estimates from the model. However, in cases where low ability workers search for new occupations, this produces a negative bias in the *across* occupation returns to experience. Although these workers will find better matches in other occupations, they will on average be lower ability since past matches are correlated with future matches. This negative selection masks the true transferability of accumulated human capital across occupations. Correctly accounting for this selection implies that skills may be more transferable than previously thought.

A number of recent papers have aimed at estimating correlated learning models in various contexts. The most closely related papers are those by Antonovics and Golan (2011) and Sanders (2010) who estimate occupational choice models in a framework of skill uncertainty. In these models the economy consist of two skills, manual and cognitive, and occupations use these skills at varying rates. These skills are assumed to be independent, but since they are used in all occupations, occupational productivities are correlated. Following Yamaguchi (2010) they derive the rate of skill use from the Dictionary of Occupational Titles to calibrate the occupational productivity correlation matrix. In this paper, the correlation matrix is estimated from the structure of wages.

This paper is organized into seven main sections. Section 2 discusses the theoretical correlated learning model. Section 3 details how the model is incorporated into an empirical framework. Section 4 outlines estimation. Section 5 describes the cohort of the NLSY97 that is used in the analysis and provides descriptive statistics on the occupational choices of these young workers. Section 6 presents the results from estimation, analyzing the role of sorting on ability for mobility patterns observed in the data. Finally, Section 7 concludes.

2 Model

This section outlines the theoretical dynamic learning model. Following the endogenous human capital and occupational choice literature of Keane and Wolpin (1997), an individual's career is defined as a sequence of discrete choices. In each period, individuals choose among $J + 3$ mutually exclusive activities, including unemployment (home production), schooling, military, or employment in one of J occupations. The first period corresponds to age 16 and each period t represents a full academic year (September to August). Let $d_{it} = c$ indicate that agent i makes choice c in period t from the choice set $C = \{\mathrm{u}, \mathrm{s}, \mathrm{m}, 1, 2, \cdots, J\}$, which follows the ordering above.

If an individual is employed in occupation j, they are paid their occupation specific productivity. Three components contribute to a workers productivity in occupation j: accumulated human capital, innate productivity (ability) in occupation j, and a transitory productivity shock. Human capital is accumulated through education and occupation specific experience. Let $\mathbf{s}_{it}$ indicate the individual's highest grade of education completed as of date t and $\mathbf{x}_{it}(j)$ be their years of experience in occupation j. The log-wage for worker i in occupation j in period t is defined as:

$$w_{itj} = \mathbf{s}_{it}\theta_j^{\mathrm{s}} + \mathbf{x}_{it}\theta_j^{\mathrm{x}} + \mu_i(j) + \eta_{itj} \tag{1}$$

The returns to education and experience are different for each occupation and are represented by θ_j^{s} and θ_j^{x}. $\mu_i(j)$ is the individual's match in occupation j and represents their innate ability, which persists over time. The $J \times 1$ vector μ_i is the individual's complete set of innate productivities for all occupations in the economy. Finally, η is a transitory shock to productivity that is independent over time and normally distributed mean zero with occupation specific variance (i.e. $\eta_{itj} \sim \mathcal{N}(0, \sigma_j^2)$).

The fundamental aspect of the agent's occupational choice decision is that in the beginning of their career, they face uncertainty over their vector of innate productivities, μ_i. Because this is a persistent component of future wages, workers have an incentive early in their career to discover where their comparative advantage lies. In order to discover this information, workers search across occupations.

It is assumed that workers are only able to acquire information about these matches through direct occupation specific experience. In each period of employment in occupation j, the worker observes his wage. The wage residual, once the part due to education and experience is removed, represents a noisy signal of their innate productivity, $\mu_i(j)$, denoted by $z_{it}(j)$. That is, workers cannot separately distinguish between their unknown ability and the technology shock, η, only observing:

$$z_{it}(j) = I[d_{it} = j](w_{itj} - \mathbf{s}_{it}\theta_j^{\mathrm{s}} - \mathbf{x}_{it}\theta_j^{\mathrm{x}})$$

Using these signals, workers are Bayesian updaters in the spirit of Miller (1984) and Jovanovic (1979). Initial beliefs are distributed multivariate normal with individual specific mean γ_i and precision matrix Δ, i.e. $\mu_i \sim \mathcal{N}(\gamma_i, \Delta)$. Given the distribution of the technology shocks, workers know the distribution of the observed signals, $z_{it}(j) \sim \mathcal{N}(\mu_i(j), \sigma_j^2)$. Workers formulate beliefs by applying Bayes' rule to a multivariate normal prior and normally distributed signals. Since the normal distribution is a conjugate distribution, the posterior beliefs will be multivariate normal as well.

Let $b_{it} = E_t[\mu_i]$ be the worker's expected value of their match vector given information at date t. This is partially determined by $\mathbf{x}_{it}$, their vector of experience, and $\overline{z}_{it}$, their average vector of signals defined as:

$$\overline{z}_{it}(j) = \begin{cases} \dfrac{1}{\mathbf{x}_{it}(j)} \displaystyle\sum_{t'=1}^{t} z_{it'}(j) & \text{if } \mathbf{x}_{it}(j) > 0 \\ 0 & \text{otherwise} \end{cases}$$

In the multivariate normal case, Bayes' rule can be written as:

$$b_{it} = \gamma_i + \zeta(\mathbf{x}_{it}, \Delta, \Sigma)[\overline{z}_{it} - \gamma_i] \tag{2}$$

Where $\zeta(\mathbf{x}_{it}, \Delta, \Sigma)$ is weighting matrix defined as:[4]

$$\zeta(\mathbf{x}_{it}, \Delta, \Sigma) = (\Delta^{-1} + \text{diag}(\mathbf{x}_{it})\Sigma^{-1})^{-1}\text{diag}(\mathbf{x}_{it})\Sigma^{-1}$$

This formulation demonstrates two intuitive aspects of the learning process. First, beliefs are formed by the deviation of the average signal form the initial prior. The amount and direction of the revision of beliefs is dictated by the weighting matrix ζ. Second, and more important to the workers' problem, is that even if a worker only has experience in one occupation, this affects their beliefs in all occupations because the off-diagonal elements of ζ are potentially non-zero.

Workers also update the precision of their beliefs. Similar to the mean, the correlation of abilities implies that experience in each occupation will affect the precision of beliefs in other occupations. Unlike the mean, the

[4] Σ is $J \times J$ diagonal matrix defined as,

$$\Sigma(j, j') = \left\{ \begin{array}{ll} \sigma_j^2 & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{array} \right\}$$

precision is only a function of the number of signals not the value of the signals. This covariance matrix is represented by $\delta_{it} = E_t[\mu_i \mu_i'] - E_t[\mu_i]E_t[\mu_i]'$ where:

$$\delta_{it} = (\Delta^{-1} + \mathrm{diag}(\mathbf{x}_{it})\Sigma^{-1})^{-1}$$

The optimal search strategy comes about as the solution to a dynamic utility maximization problem, where agents make choices to maximize the expected discounted sum of current and future utility. For each choice in the current period, workers receive a flow of utility, u_{it}^d, and if they choose one of the occupations, they receive a signal of their occupation specific productivity, $z_{it}(j)$. Although the signal has no affect on utility in the current period, it has a large impact on future utility through the evolution of beliefs. Workers may trade-off current period utility in exchange for higher expected future utility if the information value of a signal in a particular occupation is high enough.

Let $S_{it} = \{\mathbf{s}_{it}, \mathbf{x}_{it}, b_{it}, \delta_{it}, X_{it}\}$ contain all exogenous and endogenous elements affecting the worker's decision. All elements are defined above except X_{it}, which represents additional observed factors that influence occupational choices beyond those included in wages, (e.g. choices in previous periods). The per-period utility function for each choice $d \in \{\mathrm{u}, \mathrm{s}, \mathrm{m}, 1, \cdots, J\}$ is defined as:

$$u_{it}^d(S_{it}, \varepsilon_{it}) = \begin{cases} \varepsilon_{it}(d) & \text{, if } d_{it} = \mathrm{u} \\ \nu_d(S_{it}, \alpha_d) + \varepsilon_{it}(d) & \text{, if } d_{it} \in \{\mathrm{s}, \mathrm{m}\} \\ \nu_d(S_{it}, \alpha_d) + E[w_{itd} | \mathbf{s}_{it}, \mathbf{x}_{it}, b_{it}(d)] + \varepsilon_{it}(d) & \text{, if } d_{it} = 1, \ldots, J \end{cases}$$

Where ν_d represents a general function defining the mean non-pecuniary aspects of utility for choice d relative to the value of unemployment parameterized by α_d. For the employment choices, workers also derive utility from their expected occupation specific wage, which is a function of their beliefs in the current period. Finally, ε is a contemporaneous vector of random utility shocks revealed in period t, but not included in S_{it}.

Beliefs in the model evolve endogenously and stochastically. Conditional on choosing occupation j, the distribution of possible signals is given by the first and second moments of their beliefs, b_{it} and δ_{it}. Because workers are forward looking, they not only consider how their choice affects their information set next period but all future periods as well. Given a discount of future utility β and taking expectations with respect to all possible sequences of beliefs and future utility shocks, the maximal expected discounted present value V is described by,

$$V_t(S_{it}, \varepsilon_{it}) = \max_{\{d_{it'}\}_{t'=t}^{\overline{T}}} E\left\{ \sum_{t'=t}^{\overline{T}} \beta^{t'-t} u_{it'}^{d}(S_{it'}, \varepsilon_{it'}) \middle| S_{it}, \varepsilon_{it} \right\} \tag{3}$$

And $V_{\overline{T}+1} = 0$ for a terminal period $\overline{T}$.

With this expression for the value function, the optimal search strategy in the current period is,

$$d_{it} = \underset{d \in C}{\operatorname{argmax}} \left\{ u_{it}^{d}(S_{it}, \varepsilon_{it}) + \beta E[V_{t+1}(S_{it+1}, \varepsilon_{it+1}) | S_{it}, d] \right\} \tag{4}$$

3 Empirical Model

There are two central challenges for incorporating this model into an empirical framework. The first is that according to eq. (4), beliefs are an important factor influencing occupational choices. Since these beliefs are not observed directly, this creates a selection on unobservables problem. The second issue is that including more than a few occupations results in an extremely large state space. This makes it impossible to solve the dynamic programming problem in eq. (3) necessary to form the choice probabilities for estimation.

The empirical strategy for dealing with the selection problem is to form the likelihood of the observed data conditional on beliefs and then integrate over the unobservables. This section is devoted to describing how the workers' dynamic search strategy is modeled in estimation. Rather than explicitly solving for the optimal policy functions, these functions are approximated in a very flexible way and estimated from the data. This section will first describe the likelihood function and then discuss how the reduced form policy functions are identified in this dynamic framework.

The model can be estimated with a panel of data on occupational choices and wage outcomes. For each individual $i = 1, 2, \ldots, N$, we observe two sets of outcomes: their choices, $d_i = \{d_{i1}, d_{i2}, \cdots, d_{iT_i}\}$ and if they are employed their wages, $w_i = \{w_{i1}, w_{i2}, \cdots, w_{iT_i}\}$, where T_i denotes the number of periods individual i is in the sample. These outcomes are determined by both observed and unobserved variables in the state vector, S_{it}, random utility shocks, ε, occupation specific productivities, μ, and technology shocks, η.

Included in the state vector are the individual's beliefs, which are unobserved for several reasons. First we do not have consistent estimates of the wage parameters to form the wage residuals that act as the individuals' signals, $z_{it}(j)$. Second, we do not know their initial priors, γ_i, which determines if the signals are above or below their initial expectations. Finally, we do not know the weighting matrix, $\zeta(\Delta, \Sigma)$ that they use to update their beliefs and form expectations about future signals.

Taking these elements of beliefs as given and making an assumption on the distribution of ε, we can form the choice probabilities as:

$$\Pr(c|\gamma_i, \theta, \zeta(\Delta, \Sigma), \alpha) = \int_\varepsilon \left(\underset{d \in C}{\operatorname{argmax}} \left\{ u_{it}^d(S_{it}, \varepsilon) + \beta E[V_{t+1}(S_{it+1}, \varepsilon_{it+1})|S_{it}, d] \right\} = c \right) dP(\varepsilon) \tag{5}$$

The likelihood of the observed choices is:

$$L_c(d_i|\gamma_i, \theta, \zeta(\Delta, \Sigma), \alpha) = \prod_{t=1}^{T_i} \prod_{c \in C} \Pr(c|\gamma_i, \theta, \zeta(\Delta, \Sigma), \alpha)^{d_{it}=c}$$

Uncertainty in the wage equation is driven by the unobserved productivities, μ, and the unobserved technology shock, η. We have assumed that η is normally distributed, so the probability of an observed wage conditional on μ has a closed form expression:

$$\Pr(w_{it}|\theta_j, \sigma_j, \mu_i(j))^{d_{it}=j} =$$
$$\left(\frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(w_{it} - \mathbf{s}_{it}\theta_j^{\mathrm{s}} - \mathbf{x}_{it}\theta_j^{\mathrm{x}} - \mu_i(j))^2}{2\sigma_j^2}\right)\right)^{d_{it}=j}$$

The likelihood of their joint observation of wages is:

$$L_w(w_i|\theta, \sigma, \mu_i) = \prod_{t=1}^{T_i}\prod_{j=1}^{J} \Pr(w_{it}|\theta_j, \sigma_j, \mu_i(j))^{d_{it}=j}$$

With these components, the integrated likelihood of the joint probability of observed choices and wages follows:

$$L_i = \int_{\gamma} L_c(d_i|\gamma, \theta, \zeta(\Delta, \Sigma), \alpha) \int_{\mu} L_w(w_i|\theta, \sigma, \mu)\, f(\mu|\gamma, \Delta)\, d\mu\, dP(\gamma) \tag{6}$$

Where $f(\mu|\cdot)$ is the probability density function over a multivariate normal distribution and $dP(\gamma)$ is the integration over the distribution of unobserved priors. The assumption that workers have rational expectations has several implications for estimation. Not only is γ the econometric mean of wages, but it is also the worker's initial prior. Similarly, Δ not only governs the econometric distribution of μ, but it is also essential in how workers form beliefs through ζ.

The rational expectations assumption delivers an important result. Instead of solving the dynamic programming problem in eq. (5), we could approximate these choice probabilities with a reduced form policy function of the observed and unobserved variables if we had consistent estimates of θ, γ, Δ, and Σ. The rational expectations assumption implies that these parameters are identified by other features of the data, specifically the static wage equations. Jointly estimating wages and these policy functions and integrating over the unobservables correctly accounts for dynamic sorting and yields consistent estimates of the policy functions and the other parameters in the model. This quasi-structural approach is similar to Bernal and Keane

(2010).

The quasi-structural approach has the clear advantage in that it does not require the onerous task of solving the dynamic programming problem of this very large state space problem nor impose the requisite assumptions to formulate the dynamic problem. An additional benefit to this approach is that it enables us to also relax some of the previous assumptions over how workers form beliefs. Specifically, rather than parameterize the weight matrix ζ in eq. (2) with Δ and Σ, we can incorporate the belief updating rules used by the workers non-parametrically. Let $\zeta^{j}_{j',t'}$ be a parameter that represents the worker's revision of beliefs about their match in occupation j given t' signals of their match in occupation j'. Then beliefs can be represented as:

$$b_{it}(j) = \gamma_i(j) + \sum_{t'=1}^{t}\sum_{j'=1}^{J} \zeta^{j}_{j',t'} I[\mathbf{x}_{it}(j') = t'](\overline{z}_{it}(j') - \gamma_i(j'))$$

For the policy functions, what matters is how beliefs affect choices. Once we include beliefs into the policy functions, the parameters $\zeta^{j}_{j',t'}$ will be absorbed into the parameters of the policy functions. Therefore, instead of including beliefs as the state variables, we can account for their impact on choices by directly including $I[\mathbf{x}_{it}(j') = t'](\overline{z}_{it}(j') - \gamma_i(j'))$ for all $t' \in \overline{T}$ and all $j' \in J$ as state variables in the policy functions. Consequently, identification of the policy functions is driven by the rational expectations assumption on γ only and the assumption that the average signal in each occupation is a sufficient statistic in the formation of beliefs. This adds more flexibility to the model where workers can update beliefs in a way that is unknown to the econometrician, but accounted for in estimation.

The policy functions are modeled as a universal logit, with all characteristics entering all choice equations. The variables used in the approximation are defined by $\widetilde{X}_{it}(\theta, \gamma)$, where the conditioning on θ and γ indicates that this vector is dependent on other features of the model. These variables are enumerated in Appendix A. The policy functions are defined as:

$$\Omega(c|\tilde{X}_{it}(\theta,\gamma)) = \begin{cases} \dfrac{1}{1+\sum_{c'\in\{\mathrm{s},\mathrm{m},1,\ldots,J\}}\exp(\widetilde{X}_{it}\omega_{c'})} & \text{if } c=\mathrm{u} \\ \dfrac{\exp(\widetilde{X}_{it}\omega_{c})}{1+\sum_{c'\in\{\mathrm{s},\mathrm{m},1,\ldots,J\}}\exp(\widetilde{X}_{it}\omega_{c'})} & \text{if } c=\{\mathrm{s},\mathrm{m},1,\ldots,J\} \end{cases}$$

Where $\Omega(c|\cdot) \approx \Pr(c|\cdot)$, unemployment is the omitted choice, and ω are the policy function parameters.

Before discussing estimation, an important result of the productivity wage assumption can be seen in eq. (6). If individuals have homogeneous priors, then the log-likelihood in eq. (6) is additively separable in the choices and wages.[5] In this case, the wage parameters and the distributional parameters, γ and Δ, can all be consistently estimated ignoring the choice data and more importantly without solving the dynamic programming problem in eq. (3). While homogenous priors is not assumed in this paper, this result demonstrates the theoretical economic assumptions in which the observed outcome variables can be estimated as a correlated random effect, without modeling choices.[6]

4 Estimation

To accommodate initial priors, γ_i is represented by a finite mixture distribution over K types, with values, γ_k. In this way, initial priors are modeled in the same way as unobserved heterogeneity in perfect information models of occupational choice, e.g. Keane and Wolpin (1997). This level of hetero-

[5]The likelihood function becomes:

$$\ln(L_i) = \ln(L_c(d_i|\gamma,\theta,\zeta(\Delta,\Sigma),\alpha)) + \ln\left(\int_{\mu} L_w(w_i|\theta,\sigma,\mu)\, f(\mu|\gamma,\Delta)\, d\mu\right)$$

[6]Such a model can be solved using a number of standard methods, either framing the problem as a system of seemingly unrelated regressions (SUR) (Greene (2002)) or setting up a state space model and using a Kalman filter (Hamilton (1994)). One particularly useful application of this result is in problems with high-dimensional choice sets with 100's or 1,000's of choices, where modeling these choices can effectively be ignored or at least estimated independently from the other model parameters.

geneity will have a strong effect on workers' choices as well as their mean wages in all occupations. Since initial priors play an important role in the model, we allow the probability that an individual is type k to be conditional on observed data, g_i, with parameters, ϕ, defined as $\pi_k(g_i, \phi)$.[7]

The final log-likelihood for estimation is:

$$\hat{\Psi} = \underset{\Psi}{\operatorname{argmax}}$$
$$\sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k(g_i, \phi) L_c(d_i | \widetilde{X}_{it}(\theta, \gamma_k), \omega) \int_{\mu} L_w(w_i | \mu, \theta, \sigma) f(\mu | \gamma_k, \Delta) d\mu \right) \tag{7}$$

Where $\Psi = \{\phi, \gamma_k, \theta, \omega, \sigma, \Delta\}$

Although many difficult aspects of the empirical model are circumvented by the quasi-structural approach, the objective function in eq. (7) is still too difficult if not impossible to optimize. The multiple integration and the presence of the wage parameters in the choice equation make it extremely difficult to derive the analytical gradient of the objective function.[8] Because we cannot derive the analytical gradient, we would need to rely on optimizers that either compute numerical approximations to the gradient or are gradient free. Either of these methods are computationally inefficient with more than a few parameters. This is problematic because the model contains 1,350 parameters for the policy functions, 182 wage parameters, and 121 parameters governing the distribution of the unobservables. A secondary issue in estimation is that we want full flexibility in the correlation structure of μ. This means that the off-diagonal elements of Δ are non-zero, so we must constrain this matrix to be positive-definite, which may be difficult to enforce in any optimization routine.

[7] In the empirical model, g_i includes mother's highest grade completed and the highest grade of the individual at age 16. $g_i = [1\ ,\ (\textit{mother's hgc})\ ,\ I \cdot (\textit{mother's hgc missing})\ ,\ I \cdot (\textit{HGC age 16} < \textit{10})]$. $\pi_k(g_i, \phi)$ is modeled as a logistic regression, with $k = 1$ as the omitted outcome.

[8] The normality assumption on μ means that the integration over L_w with respect to μ has a closed form (although complicated) expression. So the difficulty is not primarily in the integration, but in taking the derivative of the integrated value.

The model is estimated using the Generalized Expectation and Maximization (GEM) algorithm outlined in James (2012). This approach overcomes all of the computational challenges associated with this model. First, it does not require the derivation or approximation of the gradient of the original likelihood function and can therefore accommodate an enormous parameter space with little effect on computation time. This is the main result in James (2012), which makes it possible to estimate complicated models of the type presented in this paper. Such models were previously not within reach of empirical researchers. Second, it simplifies the integration by breaking the multidimensional integration into a set of single dimensional integrals, which have simple closed form expressions. Third, it guarantees that the covariance matrix Δ is positive definite (see. Train (2007)).

Instead of directly optimizing the log-likelihood $L(\Psi)$, the GEM algorithm forms an augmented data likelihood $Q(\Psi|\Psi^m)$, which is a lower bound of the objective function at parameters Ψ^m. The crucial characteristic of $Q(\cdot)$ is that $L(\Psi^m) = Q(\Psi^m|\Psi^m)$. Therefore finding $\Psi^* = \underset{\Psi}{\operatorname{argmax}}\, Q(\Psi|\Psi^m)$ guarantees that $L(\Psi^*) > L(\Psi^m)$. This insight implies that it is possible to find values in the support of Ψ that achieve higher realizations of the likelihood $L(\cdot)$ not by maximizing $L(\cdot)$ directly, which is extremely difficult, but instead by successively maximizing $Q(\cdot)$, which is very simple. The ease of maximizing $Q(\cdot)$ makes the entire algorithm very fast, especially compared to conventional methods that rely on numerically approximating the gradient. The complete algorithm is detailed in Appendix B.

5 Data Extract and Summary: NLSY97

The model is estimated using the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97). The NLSY97 is a nationally representative sample of men and women born between 1980 and 1984. These individuals were between the ages of 12 and 18 at the time of their first interview in 1997 and interviewed annually thereafter. Round 1 to round 12 is used for the analysis, where the respondents are between the ages of 23 and 29 in

Table 1: Occupation Categories

Occ. Cat.	2002 3-Digit Census Code	Description
MGR	0010-0430	**Management occupations**
BUS	0500-0950	**Business and financial operations occupations**
PROF1	1000-1960	**Professional and related occupations: Aggregation 1**
		1000-1240 Computer and mathematical occupations
		1300-1560 Architecture and engineering occupations
		1600-1960 Life, physical, and social science occupations
PROF2	2000-3540	**Professional and related occupations: Aggregation 2**
		2000-2060 Community and social services occupations
		2100-2150 Legal occupations
		2200-2550 Education, training, and library occupations
		2600-2960 Arts, design, entertainment, sports, and media occupations
		3000-3540 Healthcare practitioner and technical occupations
SERV1	3600-3950;	**Service Occupations: Aggregation 1**
		3600-3650 Healthcare support occupations
		3700-3950 Protective service occupations
FOOD	4000-4160	**Food preparation and serving related occupations**
SERV2	4200-4650;	**Service Occupations: Aggregation 2**
		4200-4250 Building and grounds cleaning and maintenance occupations
		4300-4650 Personal care and service occupations
SALES	4700-4960	**Sales and related occupations**
OFFICE	5000-5930	**Office and administrative support occupations**
CONST	6000-6940	**Construction and extraction occupations**
		6000-6130 Farming, fishing, and forestry occupations
		6200-6940 Construction and extraction occupations
MAINT	7000-7620	**Installation, maintenance, and repair occupations**
PROD	7700-8960	**Production occupations**
TRAN	9000-9750	**Transportation and material moving occupations**

round 12.

Based on the responses to the retrospective education and employment questions, each year an individual is assigned to one of 16 mutually exclusive activities: unemployment, schooling, military service, or fulltime employment in one of 13 occupations, which are defined by their 1-digit census occupation classification code in Table 1. The assignment of activities and the population studied are described in Appendix C. The analysis focuses on the white male subsample, which includes 2,117 individuals with 16,371 person-year observations.

Table 2 summarizes the distribution of activities by age, as well as the

Table 2: Distribution of Choices By Age (%)

Age	Un-emp	Sch-ool	Mili-tary	Mgr	Bus	Prof 1	Prof 2	Serv 1	Food	Serv 2	Sales	Off-ice	Cons-truc	Main-ten	Prod	Tran
16	2	97	–	–	–	–	–	–	–	–	–	–	–	–	–	–
17	6	89	–	–	–	–	–	–	–	–	–	–	–	–	–	–
18	16	59	2	–	–	–	–	–	2	–	2	–	5	2	3	4
19	19	43	4	–	–	–	–	–	4	–	3	3	7	3	4	5
20	18	41	4	–	–	–	–	–	4	–	4	3	8	2	4	5
21	19	36	4	–	–	–	–	–	4	–	4	4	9	4	4	6
22	20	24	3	2	–	4	2	–	5	–	6	5	9	4	5	6
23	18	16	3	3	3	5	4	3	4	–	7	7	10	5	5	7
24	18	13	3	3	3	7	4	3	3	–	9	7	10	5	6	6
25	19	11	3	3	4	7	4	–	3	–	10	6	9	6	5	5
26	19	12	2	5	4	7	5	2	3	–	8	6	10	5	6	4
27	19	11	2	6	6	6	4	3	2	–	9	5	11	7	4	3
28	21	6	–	5	5	6	4	2	2	–	15	5	13	6	5	–
Obs.[b]	2407	7736	404	213	176	366	249	206	463	182	682	554	1041	477	545	670

[–] Cell less than 2%. Not reported.

[b] Number of Observations by age are: 16(2117) 17(1999) 18(1871) 19(1746) 20(1611) 21(1490) 22(1379) 23(1294) 24(1127) 25(808) 26(536) 27(298) 28(95).

number of observations at each age for this cohort. As expected, labor market attachment is weak for these young workers with about 20% neither in school or working fulltime. The share of employment is not even across the 13 occupations. In particular very few individuals in their early 20's work in any of the Manager, Business, or Professional occupations, however by their late 20's these represent about 25% of those employed.

Table 3 provides details on the link between education and occupation. The first column shows the average year of education completed for all observations of workers in these occupations for this cohort. This ranges from a low of 11.6 years for Construction occupations to a high of 15.7 years for Business and Financial Services occupations. Of the occupations traditionally grouped as high skill (Manager, Business, and Professional occupations), Manger occupations have the lowest mean education level, with on average one year less than the other three high skill occupations.

The remaining columns in Table 3 show the share of entry level occupations as well as the share of individuals who have ever worked in each occupation (conditional on entering the labor market) as of age 28 for those with at most a high school degree, those with some college, and those with a college degree or more. For high school graduates, 75% of their entry level jobs are represented by five of the 13 occupation categories, including Construction, Transportation, Food Service, Production, and Sales. Virtually none begin their career in high skill occupations. The next column shows the fraction of these workers who ever work in each occupation. Notably, more than one-third of these workers spend time in Construction occupations. Also, about 6% gain experience as a Manager, the most likely high skill occupation for these workers.

Turning to the occupation experiences of those with a college degree, more than half begin their career in either Manager, Business, or Professional occupations. Also, almost 25% will have some Sales experience by the age of 28. Finally, the patterns of those with some college are a mix between high school graduates and college graduates. However, this group is more likely to hold initial occupations in Food, Sales, and Office compared to the

Table 3: Occupations and Education

	Mean Years Education	HS Grad. or Less Share Entry[a] (%)	HS Grad. or Less Share Ever[b] (%)	Some College Share Entry (%)	Some College Share Ever (%)	College Grad. or More Share Entry (%)	College Grad. or More Share Ever (%)
MGR	14.3	–	6	3	7	7	14
BUS	15.7	–	2	2	5	11	17
PROF1	15.4	–	4	5	13	21	25
PROF2	15.3	–	4	4	10	14	18
SERV1	13.0	–	5	3	7	4	5
FOOD	11.8	13	18	16	20	3	4
SERV2	11.7	6	10	3	4	–	–
SALES	13.2	10	19	15	27	14	22
OFFICE	13.2	8	18	15	25	13	17
CONST	11.6	22	37	11	15	2	3
MAINT	12.2	7	17	7	11	–	3
PROD	11.9	12	22	5	11	4	5
TRAN	11.7	17	31	9	13	5	5

[a] Fraction of first occupation choice. Sums to one. Cells less than 2% not reported.

[b] Fraction of individuals who have entered the labor market who have ever worked in occupation. Does not sum to one. Cells less than 2% not reported.

other education groups.

6 Results

The goal of this paper is to investigate the importance of ability sorting for observed occupational choices and wages. The results section is divided into three parts. First we begin with the estimates for the distributional parameters of unobserved abilities μ. Then we move toward studying the role of sorting on these unobserved variables for career outcomes. Finally, we compare the estimates of the wage parameters from the model to OLS estimates in order to study the bias that arises when we do not control for the selection generated by the correlated learning framework. The parameter estimates not reported in the main body of the paper are presented in Appendix D.

6.1 Distribution of Abilities

The distribution of abilities, μ_i, is characterized by the individual specific initial priors γ_k and the covariance matrix Δ. The finite mixture of unobserved priors is estimated with two types (i.e. $K = 2$). The estimated Type 1's account for 60% of the population and Type 2's represent 40%. The estimates indicate that the individual type heterogeneity is highly correlated with education: Type 1's on average complete 12.6 years of education and Type 2's complete 16.1 years. Since the type estimates are so correlated with education, going forward in the description of the results it is simpler to split the population into high school graduates and college graduates.

The estimates of the distributional parameters of unobserved ability are presented in Table 4. The first line shows the differences in initial priors and the statistical significance. Here the priors are differenced against the priors of Type 1, i.e. $\gamma_2 - \gamma_1$. The priors are the individual type specific constant in the log wage equation, so the differences are interpreted as the percentage difference in initial expected wages. These estimates suggest that Type 2's have on average higher ability than Type 1's across the 13 occupations, where the difference between them is larger than zero for 10 out of the 13. Many of these differences are large and significant. The largest is in Business and Financial Services occupations where Type 2's expect 38% higher wages than Type 1's upon entering the labor market, apart from the fact that they have on average different levels of education. Conversely, Type 1's have a signifiant advantage in SERV1 occupations, which includes Healthcare Support and Protective Services. Construction occupations show the smallest difference with almost no distinction in priors across the two types.

The remaining sections of Table 4 show the estimated residual distribution of abilities across the 13 occupations after controlling for initial priors. The second row reports the standard deviation of innate productivity in each occupation and the bottom section reports the correlation structure. The standard deviation of ability refers to the *unconditional* distribution of ability in these occupations in the population, not the *conditional* distribution of ability for those that select into these occupations. The variance of occupation specific abilities is large, even after controlling for the initial level of heterogeneity. If workers have rational expectations, their 95% confidence interval of their uncertainty for their Manager ability (for example) is that their productivity will be either 57% lower than they initially expected, or 57% higher than they expected. Put another way, in the population the top mangers are 57% more productive than the average manager, while the least productive would be 57% less productive than average. These variances differ considerably across occupations, with the largest variances among Sales, Manger, and Business and the lowest variances in Service and Transportation occupations.

Occupation specific abilities are highly correlated across occupations. In general, these values are all positive, with only one negative (not significant) correlation between PROF2 (which includes Social Service occupations and education related occupations) and Business and Financial Services occupations. The fact that the correlations are mostly positive implies that there is a general level of productivity that persists even after controlling for initial

Table 4: Distribution of Abilities

	Mgr	Bus	Prof 1	Prof 2	Serv 1	Food	Serv 2	Sales	Off-ice	Cons-truc	Main-ten	Prod	Tran
Difference in Initial Priors													
$\gamma_2 - \gamma_1$	0.23	0.38^{*}	0.15	-0.07	-0.13^{**}	0.17^{*}	0.33^{***}	0.17	0.28^{**}	0.01	0.11	0.22	-0.07
Standard Deviation of Ability													
$\mathrm{std}(\mu_i(j))$	0.29	0.29	0.25	0.28	0.22	0.24	0.22	0.31	0.24	0.29	0.28	0.26	0.23
Correlation													
MGR	1.00	–	–	–	–	–	–	–	–	–	–	–	–
BUS	0.25	1.00	–	–	–	–	–	–	–	–	–	–	–
PROF1	0.83^{***}	0.13	1.00	–	–	–	–	–	–	–	–	–	–
PROF2	0.75^{**}	-0.37	0.76^{***}	1.00	–	–	–	–	–	–	–	–	–
SERV1	0.93^{***}	0.22	0.88^{***}	0.81^{**}	1.00	–	–	–	–	–	–	–	–
FOOD	0.70^{**}	0.60	0.68^{**}	0.22	0.60^{**}	1.00	–	–	–	–	–	–	–
SERV2	0.72^{***}	0.32	0.77^{***}	0.64^{**}	0.91^{***}	0.41^{*}	1.00	–	–	–	–	–	–
SALES	0.57^{***}	0.40	0.44^{*}	0.36	0.73^{***}	0.30^{*}	0.85^{***}	1.00	–	–	–	–	–
OFFICE	0.39^{*}	0.75^{**}	0.55^{***}	0.00	0.53^{**}	0.57^{**}	0.72^{***}	0.66^{***}	1.00	–	–	–	–
CONST	0.75^{***}	0.04	0.64^{**}	0.80^{*}	0.77^{***}	0.23	0.69^{***}	0.38	0.18	1.00	–	–	–
MAINT	0.74^{***}	0.38	0.86^{***}	0.52^{*}	0.86^{***}	0.66^{**}	0.88^{***}	0.77^{***}	0.81^{***}	0.42^{***}	1.00	–	–
PROD	0.83^{***}	0.26	0.62^{*}	0.44	0.63^{**}	0.70^{*}	0.36	0.35	0.29	0.39^{**}	0.58^{**}	1.00	–
TRAN	0.80^{***}	0.56	0.40^{**}	0.32	0.69^{***}	0.65^{**}	0.51^{***}	0.66^{**}	0.38	0.47^{***}	0.52^{***}	0.72^{***}	1.00

***,**,* Significantly different from zero at 1%, 5%, and 10% level respectively. Not Reported for estimates of standard deviation of ability. Constructed from 800 non-parametric bootstraps of the model.

priors. As individuals learn they are more (less) productive than they initially expected in one occupation, they will likely infer that they are better (worse) at all things than initially expected.

While occupation abilities are very correlated, important differences appear. One example is among the three service occupations, which by census definitions should all be very related. This is true between the SERV1 and SERV2 category, where the correlation coefficient is one of the highest in the table at 0.91. However, these two service categories are not nearly as correlated with Food Service, with a correlation coefficient of 0.41 between Food Service and SERV2.

6.2 Learning and Occupational Sorting

We now turn to examining the extent to which workers sort on occupation specific productivities and the extent to which this sorting is effected by the fact that abilities are correlated. The main tool for this analysis is the estimated policy functions and the primitives of the underlying ability distribution. Given these values, we can regenerate the data and study workers' responses to new information over their careers'. The focus is on sorting relative to initial expectations, measured as $\mu_i(j)-\gamma_i(j)$.[9] Since these variables relate to the individual specific constant in the log-wage equation, the value of this measure has a very clear interpretation as a percent change in wages.

To begin the analysis, we first examine the empirical hazard of workers choosing an occupation conditional on their ability. The upper plot of Figure 1 shows the average probability of working in Construction at age 28 for high school graduates whose entry occupation was Construction as a function of their Construction ability differenced by their initial prior. The bottom plot preforms a similar analysis for college graduates and Sales occupation. According to Table 3, both of these occupations represent important entry occupations for the respective education levels.

[9] Sorting on initial priors is not studied in this paper. In some ways this type of sorting resembles that of the perfect information model of Keane and Wolpin (1997), which is not surprising since this form of unobservable is modeled in a very similar way.

For both of these occupations, Figure 1 confirms our main intuition, that those who are less capable in their entry occupation are less likely and those that are more capable are more likely to continue working in their entry occupation. High school graduates beginning their career in Construction who are 40% less productive than average (about -1.5 standard deviations below the mean) have only a 25% probability of choosing Construction at age 28. By contrast, the mean probability of choosing Construction at age 28 is more than 60% for workers whose Construction ability is +1.5 standard deviations above average.

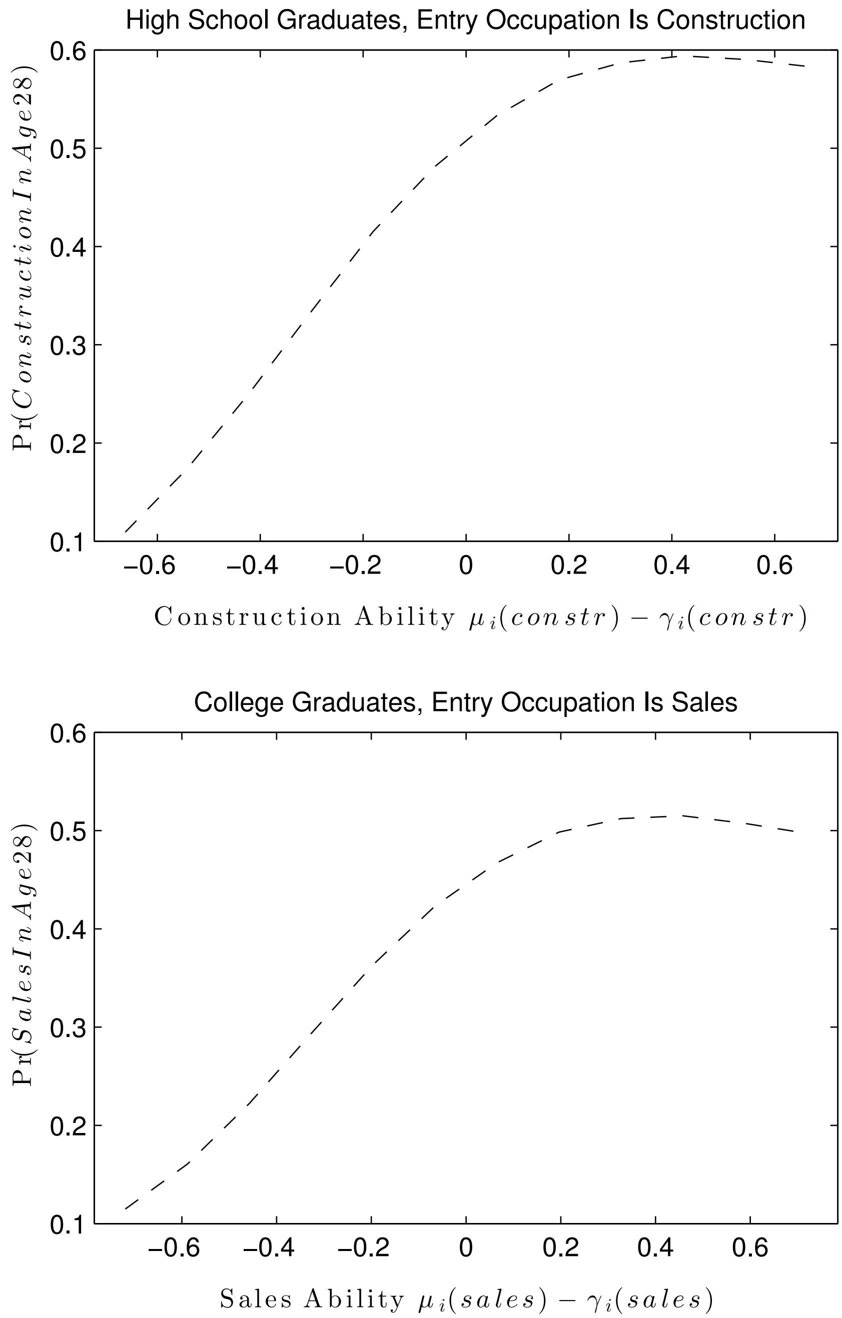
However, for both of these occupations, the hazard rate plateaus for high ability workers and even begins to decline for very high ability workers. This underscores the complicated selection problem generated by the correlated learning structure. In concordance with the matching model, low ability workers are likely to leave these occupations in pursuit of better matches in other occupations. On the other hand, high ability workers may become mobile as well, since the correlation structure may increase the value of their other employment options.

To show the effects of this sorting for Construction, the top section of Figure 2 plots the conditional distribution for high school graduates that select into Construction at three points in time. The solid line plots the unconditional distribution of abilities in Construction, which is representative of workers choosing occupations randomly or the initial sorting into Construction.[10] The dashed line plots the distribution of abilities for those that choose Construction at age 21, and the dash-dot line plots the distribution for those that select Construction at age 28. This plot provides strong evidence that workers are selecting positively on ability in Construction as the probability mass moves up and to the right with age.[11] This means that these workers are becoming on average higher ability and more concentrated around higher ability over time. The vertical dash-dot line shows

[10] Ability and productivity in this context refers to μ only, which abstracts from the fact that workers may become more productive over time through accumulated human capital.

[11] The results are even stronger if the plots are conditional on experience in Construction rather than age.

Figure 1: Hazard Rate of Staying in Entry Occupation



the average ability in construction at age 28 for high school graduates. For Construction, wages are on average 6% higher due to this positive sorting compared to a world where workers choose occupations randomly and we hold fix the distribution of accumulated human capital.

The sorting patterns in Construction are illustrative of occupations that exhibit positive sorting. An important implication of the correlated learning framework is that not all occupations will experience positive sorting. In fact, in some occupations we may observe negative sorting, where high ability workers are the most likely to leave these occupations and the average ability decreases over time. An example of average negative selection occurs for high school graduates that select into Food Service occupations. For High school graduates, Figure 2 plots how the distribution of ability in Food Service evolves with age for those that work in Food Service.

Table 5 summarizes the average sorting on ability for each occupation at age 28 for high school graduates and college graduates and compares these values against two other major contributors to wage growth: education and own occupation specific experience. Most occupations show strong evidence of positive sorting. Of the eight occupations most populated by high school graduates, Manager, Sales, Construction, and Production show an increase of about 6%, Maintenance and Transportation show a smaller increase of 3%, and Food and Office show negative values. Similar patterns and levels persist for college graduates. For most occupations, the average increase in wages that occurs through sorting on ability is more than the average increase in wages which would occur if all individuals were endowed with one additional year of education. Similarly, the average improvement in wages for many occupations is almost equivalent to the increase in wages that would occur if all workers received one additional year of experience.

Looking only at the average effect of sorting may mask some important patterns in the data. For example, the average ability in Construction at age 28 for high school graduates is 0.06. This is composed of individuals who choose Construction and discover they have high ability and stay, those that discover low ability and leave, others who may have begun their career

Figure 2: Conditional Distribution of Ability In Construction and Food Service Occupations By Age for High School Graduates

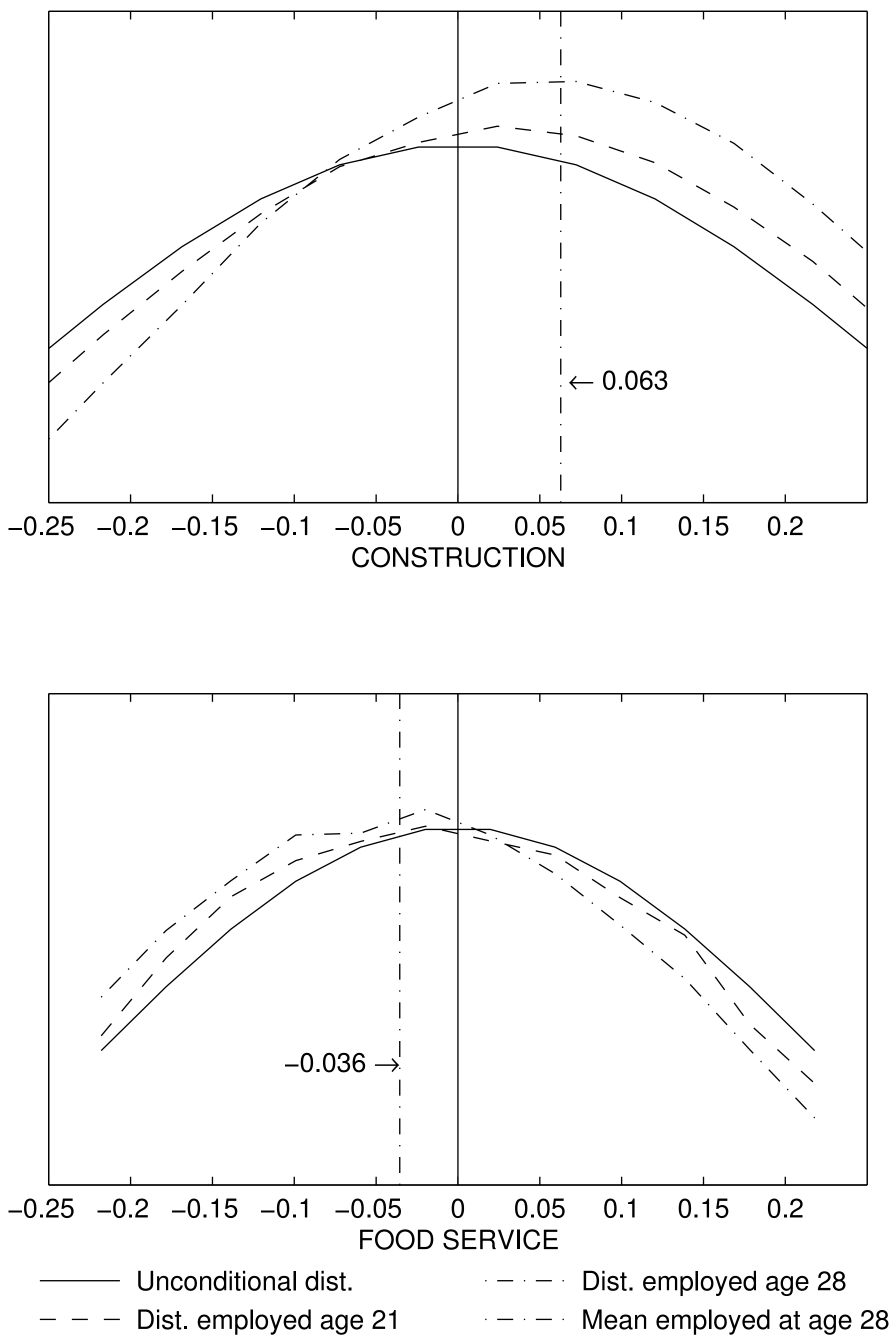


Table 5: Comparison of Wage Returns at Age 28

	Wage Return For One Year of Accum Human Capital		HS Grad[a]		Col Grad	
	Educ. θ_j^{s}	Own Expr. $\theta_j^{\mathrm{x}}(j)$	Share Emp (%)	Mean Ability Age 28[b]	Share Emp (%)	Mean Ability Age 28
MGR	0.053	0.104	4	0.067	17	0.072
BUS	0.018	0.080	–	–	16	−0.007
PROF1	0.052	0.102	–	–	14	0.056
PROF2	0.096	0.027	–	–	7	0.070
SERV1	0.074	0.056	–	–	3	0.039
FOOD	0.030	0.050	7	−0.036	–	–
SERV2	0.007	0.061	–	–	–	–
SALES	0.043	0.055	10	0.063	21	0.026
OFFICE	0.022	0.061	8	−0.004	12	0.003
CONST	0.038	0.075	24	0.062	–	–
MAINT	0.023	0.083	21	0.032	–	–
PROD	0.036	0.071	9	0.061	–	–
TRAN	0.023	0.054	9	0.033	–	–

[a] Cells with employment share less than 3% not reported.
[b] $E[\mu(j) - \gamma(j)|d_{28} = j]$, for row j

in Transportation, discovered a low match, and subsequently moved to Construction, etc. Tables 6 decomposes these averages for high school graduates in Transportation, Office, and Manager occupations as examples to further highlight important dimensions of the data.

The top section of Tables 6 provides information on high school graduates working in Transportation at age 28. Each row shows the entry occupation of these workers. The first column shows the share of age 28 Transportation workers coming from each entry occupation. For each entry occupation, the remaining columns compare the average ability in the entry occupation and the average ability in Transportation for those workers choosing Transportation relative to the other workers beginning their career in that entry occupation. Looking at transitions at this level of detail may generate some small sample issues. To address this problem, these statistics are simulated from the non-parametric bootstrap of parameters and statistical significance assigned based off of the estimated confidence interval.

Table 5 showed that sorting by Transportation workers on average accounted for 3% of wages. This is much smaller than many of the other occupations. Tables 6 demonstrates that there are large differences in this average when you compare individuals by their entry occupation. For example, 38% of Transportation workers at age 28 began their career in Transportation. For these workers choosing to stay in Transportation, sorting on average represents 8% of wages, more than double the average for all Transportation workers. This means that driving down the average is a flow into Transportation of poorly matched workers. This phenomenon is a byproduct of the correlation of abilities. As low ability workers seek out better matches by changing occupations, they will likely find on average lower, but better matches in other occupations. This is apparent in looking at the 11% of transportation workers who began their career in Construction. These workers are on average 20% less productive in Construction than the other workers whose entry occupation was Construction. The estimated correlation coefficient between Construction and Transportation is 0.5, meaning that these workers will likely have lower Transportation ability as well. This is clearly true as their average transportation ability is -0.05, which drags down the total average. A different measure of sorting that could apply to these workers is the 15% increase in wages they experience (holding other things fixed) by avoiding Construction.

The middle section of Table 6 decomposes the average ability for those working in Office and Administrative Support occupations, which had a negative value. The influx of low ability workers is the primary factor driving down the average sorting on ability observed for this occupation. Those who began their career in Office and choose Office at age 28 show no positive sorting on ability. However, low ability workers from Sales, Construction, Maintenance, and Production bring down the average over time. Importantly though, each of these workers finds a better match than their entry occupation.

There is strong evidence that workers sort into occupations where they discover that they are high ability. We are also interested in the degree that workers discovering a high ability in one occupation may use this information to sort into other occupations. The important occupation for this analysis is Manager occupations. Virtually no high school graduate begins their career as a Manager, yet 4% are in these occupations by age 28. The bottom section of Table 6 shows the career patterns for these high school graduates who are working in Manager occupations at age 28. Manager occupations show the largest level of ability sorting. This is extremely suggestive of the stepping-stone type mobility in Jovanovic and Nyarko (1997). The two largest values are for Sales occupations and Construction occupations. Construction workers that become Managers are on average 21% more productive Construction workers than the average worker beginning in Construction. Because of the correlation they are also 17% more productive Managers. Positive sorting on Manager ability is less important in other occupations, for example Transportation.

To conclude the analysis on sorting, Table 7 summarizes the aggregate effect of sorting for all individuals in the population. The analysis divides workers into two groups: those who discover they have higher ability and those who discover they have lower ability than initially expected in their

Table 6: Average Ability By Age 28 Occupation and Entry Occupation, High School Graduates

Entry Occupation	Share Age 28 Occ.(%)	Entry Occ. Match[a]	Age 28 Occ. Match[b]
Transportation at Age 28			
FOOD	7	−0.05	−0.01
SALES	7	0.03	0.03
OFFICE	15	0.09	0.06*
CONST	10	−0.20**	−0.05
MAINT	5	−0.20	−0.04
PROD	11	−0.04	−0.00
TRAN	38	0.08*	0.08*
Office and Administrative Support at Age 28			
FOOD	19	0.02	0.02
SALES	10	−0.14*	−0.05
OFFICE	21	0.01	0.01
CONST	8	−0.30***	−0.03
MAINT	5	−0.09	−0.04
PROD	10	−0.20&	−0.03
TRAN	17	−0.01	0.00
Manger at Age 28			
FOOD	11	−0.02	0.04
SALES	7	0.26*	0.16*
OFFICE	9	−0.07	0.03
CONST	15	0.21**	0.17**
MAINT	–	–	–
PROD	10	−0.13	−0.08
TRAN	24	−0.02	0.01

[a] $E[\mu(j)-\gamma(j)|d_{entry}=j,d_{28}=j']-E[\mu(j)-\gamma(j)|d_{entry}=j]$

[b] $E[\mu(j')-\gamma(j')|d_{entry}=j,d_{28}=j']-E[\mu(j')-\gamma(j')|d_{entry}=j]$

***,**,* Significantly different from zero at 1%, 5%, and 10% level respectively. Constructed from 800 non-parametric bootstraps of the model.

& Significantly different from zero at 15% level.

Table 7: Occupational Sorting at Age 28, All Workers

	Share of Total (%)	Entry Occ. Match[a]	Age 28 Occ. Match[b]	Difference[c]
ABOVE Average Entry Occupation Match: $[\mu(d_{entry}) - \gamma(d_{entry})] > 0$				
Unemployed Age 28	9	0.199	–	–
Age 28 Occ. Same as Entry	15	0.218	0.218	0.000
Age 28 Occ. Different as Entry	26	0.211	0.135	−0.076
BELOW Average Entry Occupation Match: $[\mu(d_{entry}) - \gamma(d_{entry})] < 0$				
Unemployed Age 28	12	−0.219	–	–
Age 28 Occ. Same as Entry	11	−0.177	−0.177	0.000
Age 28 Occ. Different as Entry	27	−0.220	−0.092	0.127
Total	100	0.000	0.029	–

[a] $E[\mu(d_{ent}) - \gamma(d_{ent})]$
[b] $E[\mu(d_{28}) - \gamma(d_{28})]$
[c] $E[\mu(d_{28}) - \gamma(d_{28})] - E[\mu(d_{ent}) - \gamma(d_{ent})]$

entry occupation. Each group is equally represented by definition. Each group is further broken down by those who are unemployed at age 28, those working in their entry occupation at age 28, and those working in an occupation different from their entry occupation at age 28.

The first column shows the share of each segment of the population as a percent of the total population. Those with above average ability in their entry occupation are 33% more likely to remain in their entry occupation than those discovering they are below average ability in their entry occupation. These two groups move to new occupations at the same rate, but those discovering low ability in their entry occupation are more likely to be unemployed at age 28.

For each segment of the population, the remaining columns show the average ability in their entry occupation, their average ability in their age 28 occupation, and the difference. The bottom row reports the total. By construction, the average ability of entry occupation is zero, since no workers have any information to sort on. By age 28, the average ability increases to almost 3%. From the earlier analysis, this number is the average of all occupations, with some having a 6% increase, others having a 3% increase, and some having negative values.

Using the entry match as a baseline, we can compare the match for the occupations workers choose at age 28. For workers with above average productivity in their entry occupation, they have above average productivity in their age 28 occupation, either from staying in their entry occupation or moving to a new occupation. Those changing occupations on average have lower, but still positive, ability in their new occupation. This was seen in both Construction workers and Sales workers moving to Manger occupations, where each group had high entry abilities than Manager abilities.

Workers with a below average ability in their entry occupation continue to have below average ability in their age 28 occupation on average. However, the workers who change occupations move to occupations where they are substantially more productive, a difference of 12%. These workers sorting into occupations where they have higher, but still negative, abilities are one of the main drivers of the increase in average ability in the population.

6.3 Selection Bias

The correlated learning framework generates a very interesting selection problem. For most occupations there is strong evidence that workers sort positively on ability. Failing to account for this selection will produce an upward bias in the returns to experience in these occupations. However, as low ability workers try new occupations, the correlation implies they will on average be lower ability in other occupations. Failing to account for this selection may downward bias the estimates of the across occupational returns to accumulated human capital, understating the transferability of accumulated skills across occupations.

This section studies the bias in the OLS estimates when we do not control for selection. What makes this analysis challenging is that a bias in one parameter biases all of the parameters. To address this, we compute the theoretical bias which would occur if consistent estimates of all of the

Table 8: OLS Bias in Wage Estimates[a]

	Food	Sales	Off-ice	Cons-truc	Main-ten	Prod	Tran
$\mathbf{x}_{it}$ FOOD	0.009	0.015	0.015	-0.013	-0.033	0.008	$-0.020^{\&}$
$\mathbf{x}_{it}$ SALES	-0.011	0.030^{**}	-0.020	0.009	0.004	-0.014	-0.001
$\mathbf{x}_{it}$ OFFICE	-0.107^{**}	-0.014	0.006	-0.020^{*}	-0.021	0.018	0.005
$\mathbf{x}_{it}$ CONST	$-0.033^{\&}$	-0.006	-0.017^{*}	0.014^{***}	-0.019^{**}	-0.014^{*}	-0.024^{***}
$\mathbf{x}_{it}$ MAINT	-0.028	-0.086^{*}	-0.016	-0.015^{***}	0.013^{*}	-0.007	$-0.024^{\&}$
$\mathbf{x}_{it}$ PROD	0.003	-0.018	-0.018^{*}	-0.005	0.007	0.017^{*}	-0.008
$\mathbf{x}_{it}$ TRAN	-0.074^{**}	-0.029^{*}	-0.005	0.000	-0.003	-0.003	0.019^{***}

[a] This analysis uses the model estimates to simulate choices and wages for high school graduates with Type 1 priors and computes the theoretical bias taking all of the other parameters as given.

***,**,* Significantly different from zero at 1%, 5%, and 10% level respectively. Constructed from 800 non-parametric bootstraps of the model.

$^{\&}$ Significantly different from zero at 15% level.

parameters were available except the parameter of interest. In this case the bias in the return to experience in occupation j' for wages in occupation j is defined as $cov(\mathbf{x}_{it}(j'), \mu_i(j)|d_{it} = j)/var(\mathbf{x}_{it}(j')|d_{it} = j)$. In addition, to abstract from sorting by education and type, we also condition this analysis on high school graduates with Type 1 priors.

Table 8 shows the bias in the estimated return to accumulated human capital for the seven most common occupations for high school graduates. As expected, most of the own occupation specific returns (the diagonal elements) are significantly biased upward as workers sort positively on ability. The two exceptions are Food Service and Office and Administrative support, which showed little sorting in the earlier analysis. The bias appears most significant for Sales, Construction, and Transportation. Tables 5 shows the estimated own experience return for these occupations from the model. The bias is largest in Sales of close to 50%. For Transportation the bias is 35% and almost 20% for Construction.

Looking at the bias in the across occupational returns to experience (the off-diagonal elements), for many occupations, the selection produces a downward bias. Of the 42 across occupation experience returns reported, 31 have a negative bias, with 10 being statistically significant. One of the largest and most significant is the return for Construction experience in Transportation, where Table 6 demonstrated that low ability Construction workers were likely to sort into Transportation.

7 Conclusion

This paper develops and estimates a multidimensional correlated learning model to study early career wage growth and occupational choices. By allowing occupation specific productivities to be flexibly correlated, the model facilitates both sorting through experimentation, described in Miller (1984), where workers discovering low ability in an occupation are likely to try new occupations in pursuit of a better match, and sorting through promotion, described in Jovanovic and Nyarko (1997), where workers discovering high ability in an occupation become more likely to be move to new occupations because of their high performance in their first occupation.

The correlated learning approach produces an extremely challenging selection problem as the worker's optimal search strategy is a complicated function of the unknown model parameters. Previous work has avoided this selection problem by imposing assumptions on the correlation structure, either assuming occupation specific productivities are completely independent or perfectly correlated, which imply the optimal search strategy to be imposed in estimation. In this paper, the underlying correlation structure of occupation specific productivity is a key parameter of interest, and the optimal search strategy is an endogenous feature of the model. The methodological approach used in this paper can be advantageously extended to the broader class of learning models where continuous outcomes are observed along with choices. For example, major choice and grades, pharmaceutical demand when data is available on side-effects, or health choices when measures of health are observed.

The model is estimated on the National Longitudinal Survey of Youth 1997. On average, sorting on ability has a positive and strong effect on wages. By age 28, the average increase in wages due to sorting on ability is approximately equal to the average increase in wages that would occur if all workers were endowed with an additional year of education. The results provide strong evidence that sorting occurs in multiple dimensions. Workers who discover a low ability in their current occupation are very likely to try new occupations. In most cases these workers find higher matches. On the other hand, while workers that discover they are high ability in an occupation are more likely to stay, they also become more likely to move to high skill occupations like Manager.

Finally, while the empirical model did not require estimating the structural utility parameters directly, it delivered consistent estimates of all of the wage and distributional parameters, as well as consistent estimates of the conditional choice probabilities. Taking these values as given, James (2011) shows how the structural utility parameters can be recovered in a computationally efficient second stage using the insights of Arcidiacono and Miller (2010).

References

D. Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44 (3):1007–1040, August 2003.

K. Antonovics and L. Golan. Experimentation and job choice. Working Paper, March 2011.

P. Arcidiacono and J. B. Jones. Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946, May 2003.

P. Arcidiacono and R. A. Miller. Ccp estimation of dynamic discrete choice models. Duke University, July 2010.

R. Bernal and M. Keane. Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156:164–189, 2010.

G. Crawford and M. Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, July 2005.

R. Gibbons, L. F. Katz, T. Lemieux, and D. Parent. Comparative advantage, learning, and sectoral wage determination. *Journal of Labor Economics*, 23(4), 2005.

W. H. Greene. *Econometric Analysis*. Prentice Hall, 2002.

J. D. Hamilton. State-space models. *Handbook of Econometrics*, 4, 1994.

J. James. Ability matching and occupational choice. Federal Reserve Bank of Cleveland, Oct. 2011.

J. James. A tractable estimator for general mixed multinomial logit models. Federal Reserve Bank of Cleveland, May 2012.

B. Jovanovic. Job matching and the theory of turnover. *The Journal of Political Economy*, 87(5):972–990, Oct. 1979.

B. Jovanovic and Y. Nyarko. Stepping-stone mobility. *Carnegie-Rochester Conference Series on Public Policy*, 46(289-325), 1997.

G. Kambourov and I. Manovskii. Occupational specificity of human capital. *International Economic Review*, 50(1):63–115, February 2009.

M. Keane and K. Wolpin. The career decisions of young men. *Journal of Political Economy*, 105(3):473–522, June 1997.

B. P. McCall. Occupational matching: A test of sorts. *The Journal of Political Economy*, 98(1):45–69, Feb. 1990.

R. Miller. Job matching and occupational choice. *The Journal of Political Economy*, 92(6):1086–1120, Dec. 1984.

D. Neal. The complexity of job mobility among young men. *Journal of Labor Economics*, 17(2):237–261, Apr. 1999.

T. Papgeorgiou. Learning your comparative advantage. Penn State University, Jun. 2010.

R. Pavan. Career choice and wage growth. University of Rochester, July 2009.

C. Sanders. Skill uncertainty, skill accumulation, and occupational choice. Washington University, St. Louis, December 2010.

K. L. Shaw. Occupational change, employer change, and the transferability of skills. *Southern Economic Journal*, 53(3):702–719, Jan. 1987.

K. Train. A recursive estimator for random coefficient models. *Working Paper*, 2007.

S. Yamaguchi. Tasks and heterogeneous human capital. McMaster University, June 2010.

A Conditional Choice Probabilities

Table 9: List of Variables Included in Choice Policy Functions: $\widetilde{X}_{it}(\theta,\gamma_k)$

Definition	Number of Parameters
age_{it}	1
$I[age_{it} \in \{16, 17\}]$	1
$I[age_{it} \in \{18, 19, 20, 21\}]$	1
Highest grade completed	1
High school degree	1
BA degree	1
Lag Choice	15
Ever have experience in military and each occupation	14
Cum. Experience in military and occupation	14
$I[\gamma_i = \gamma_k]\,\forall k \in K$	2
$I[\mathbf{x}_{it}(j) \geq 1](\overline{z}_{it}(j) - \gamma_k(j))\,\forall j \in J$	13
$I[\mathbf{x}_{it}(j) \geq 2](\overline{z}_{it}(j) - \gamma_k(j))\,\forall j \in J$	13
$I[\mathbf{x}_{it}(j) \geq 3](\overline{z}_{it}(j) - \gamma_k(j))\,\forall j \in J$	13
Total Each Equation	90
	$\times$15
Total Number of Parameters in Policy Functions	1,350

B Details of Estimation

The quasi-structural model is estimated using a Generalized Expectation and Maximization (GEM) algorithm outlined in James (2012). The GEM algorithm is a modified EM algorithm that uses a simple step to update the choice parameters rather than the full M-step of the EM algorithm. Rather that directly maximizing eq. (7), the GEM algorithm instead iteratively maximizes an augmented data likelihood $\Psi^{m+1} = \underset{\Psi}{\operatorname{argmax}}\, Q(\Psi|\Psi^m)$, which takes as input the previous iterations estimates Ψ^m.

Letting $f(\cdot|\xi, \Upsilon)$ denote the cdf of a multivariate normal with mean ξ

and covariance Υ. The augmented data likelihood takes the form:

$$Q(\Psi|\Psi^m) = \sum_{i=1}^{N}\sum_{k=1}^{K} q_{ik}^m \int_\mu \ln\Big(\pi_k(g_i,\phi) L_c(d_i|\widetilde{X}_{it}(\theta,\gamma_k),\omega) \\ \times L_w(w_i|\mu,\theta,\sigma) f(\mu|\gamma_k,\Delta)\Big) f(\mu|\xi_{ik}^m,\Upsilon_{ik}^m) d\mu \qquad (8)$$

The augmented data likelihood differs from the original likelihood in two ways. First it includes two additional elements, q_{ik}^m and $f(\mu|\xi_{ik}^m,\Upsilon_{ik}^m)$. These can be interpreted as probability density functions of the unobserved variables γ_i and μ conditional on the individual's data and the current parameter estimates. q is a discrete probability of the unobserved prior, and f is a multivariate normal distribution of unobserved ability, μ. For now we will take these as given and discuss the maximization of eq. (8). Later we will return to the derivation of these densities.

The second difference in the augmented data likelihood and the original likelihood is that the log function is inside of the integration. This is an extremely desirable property of the EM algorithm discussed in Arcidiacono and Jones (2003), which implies additive separability of many of the parameters and allows us to maximize the parameters independently.

The maximization of each parameter has a closed form expression, so the algorithm is extremely fast. The update of each parameter is discussed below.

Update ϕ: Parameters for initial conditions. ϕ enters eq. (8) as:

$$\operatorname*{argmax}_{\phi} \sum_{i=1}^{N}\sum_{k=1}^{K} q_{ik}^m \ln(\pi_k(g_i,\phi))$$

Where

$$\pi_k(g_i,\phi) = \begin{cases} \dfrac{1}{1+\exp(g_i\phi)} & \text{if } k=1 \\ \dfrac{\exp(g_i\phi)}{1+\exp(g_i\phi)} & \text{if } k=2 \end{cases}$$

This represents a weighted logit, whose maximization does not have a closed form solution. The GEM algorithm in James (2012) shows that we can update the parameters by computing one modified Newton-Raphson step.

$$\phi^{m+1} = \phi^m - B_\phi^{-1}\left(\sum_{i=1}^{N}\sum_{k=1}^{K} g_i(q_{ik}^m - \pi_k(g_i, \phi^m))\right)$$

Where $B_\phi^{-1} = -4\left(\sum_{i=1}^{N} g_i' g_i\right)^{-1}$

Update ω: Parameters for policy functions. ω enters eq. (8) as:

$$\underset{\omega}{\operatorname{argmax}} \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T_i}\sum_{c\in C}(d_{it}=c) q_{ik}^m \ln(\Omega(c|\tilde{X}_{it}(\theta,\gamma_k)))$$

Where

$$\Omega(c|\tilde{X}_{it}(\theta,\gamma_k)) = \begin{cases} \dfrac{1}{1+\sum_{c'\in\{\mathrm{s,m},1,\dots,J\}}\exp(\widetilde{X}_{it}\omega_{c'})} & \text{if } c = \mathrm{u} \\ \dfrac{\exp(\widetilde{X}_{it}\omega_c)}{1+\sum_{c'\in\{\mathrm{s,m},1,\dots,J\}}\exp(\widetilde{X}_{it}\omega_{c'})} & \text{if } c = \{\mathrm{s,m},1,\dots,J\} \end{cases}$$

This represents a weighted logit, so we apply the same lower bound result as the initial condition parameter. Since this is a multivariate logit, the parameters must be appropriately vectorized but has the general form

$$\omega^{m+1} = \omega^m - B_\omega^{-1}\left(\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T_i} q_{ik}^m \widetilde{X}_{it}(d_{it} - \Omega(\cdot|\tilde{X}_{it}(\theta,\gamma_k)))\right)$$

See James (2012) for the derivation of B_ω^{-1}.

Update γ_k: Initial priors, independently for each k. γ_k enters eq. (8) in both the choice equations and as the mean of the distribution. Arcidiacono and Jones (2003) show that while less efficient, we can obtain consistent estimates of the parameters by maximizing over only part of the likelihood. Ignoring that γ_k enters the choices we have:

$$\underset{\gamma_k}{\operatorname{argmax}} \sum_{i=1}^{N} q_{ik}^m \int_\mu \ln(f(\mu|\gamma_k,\Delta^m)) f(\mu|\xi_{ik}^m,\Upsilon_{ik}^m) d\mu$$

This has a closed form solution as:

$$\gamma_k^{m+1} = \frac{\sum_{i=1}^N q_{ik}^m \xi_{ik}^m}{\sum_{i=1}^N q_{ik}^m}$$

Update Δ: Covariance of abilities. Δ enters eq. (8) as:

$$\operatorname*{argmax}_{\Delta} \sum_{i=1}^{N} \sum_{k=1}^{K} q_{ik}^m \int_{\mu} \ln(f(\mu|\gamma_k^{m+1}, \Delta)) f(\mu|\xi_{ik}^m, \Upsilon_{ik}^m) d\mu$$

This has a closed form solution as,

$$\Delta^{m+1} = \sum_{k=1}^{K} \left(\frac{\sum_{i=1}^N q_{ik}^m (\Upsilon_{ik}^m + (\xi_{ik}^m)(\xi_{ik}^m)')}{N} - \frac{\sum_{i=1}^N q_{ik}^m}{N} (\gamma_k^{m+1})(\gamma_k^{m+1})' \right)$$

Update θ_j: Wage parameters for occupation j independently for each occupation. θ_j enters eq. (8) in both the choice equation and wage equation. Applying the result of Arcidiacono and Jones (2003) again and ignoring that θ enters the choices:

$$\operatorname*{argmax}_{\theta_j} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T_i} (d_{it} = j) q_{ik}^m \int_{\mu(j)} \ln(\Pr(w_{it}|x_{it}, \mu, \theta_j, \sigma_j^m)) f(\mu(j)|\xi_{ik}^m, \Upsilon_{ik}^m) d\mu$$

This is a weighted least squares problem with a closed form solution

$$\theta_j^{m+1} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} (d_{it} = j) x_{it}' x_{it} \right)^{-1} \left(\sum_{i}^{N} \sum_{t=1}^{T_i} (d_{it} = j) x_{it}' (w_{it} - E(\mu_i(j))) \right)$$

Where $E(\mu_i(j)) = \sum_{k=1}^{K} q_{ik}^m \xi_{ik}^m(j)$.

Update σ_j: Variance of technology shock independently for each occupation j. σ_j enters eq. (8) as:

$$\operatorname*{argmax}_{\theta_j} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T_i} (d_{it} = j) q_{ik}^m \int_{\mu(j)} \ln(\Pr(w_{it}|x_{it}, \mu, \theta_j^{m+1}, \sigma_j)) f(\mu(j)|\xi_{ik}^m, \Upsilon_{ik}^m) d\mu$$

This is a weighted least squares problem with a closed form solution

$$\left(\sigma_j^{m+1}\right)^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(d_{it}=j)\left((w_{it}-x_{it}\theta_j^{m+1})^2 - 2(w_{it}-x_{it}\theta_j^{m+1})E(\mu_i(j)) + E^2(\mu_i(j))\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(d_{it}=j)}$$

Where $E(\mu_i(j)) = \sum_{k=1}^{K} q_{ik}^m \xi_{ik}^m(j)$ and $E^2(\mu_i(j)) = \sum_{k=1}^{K} q_{ik}^m(\Upsilon_{ik}^m(j,j)+(\xi_{ik}^m(j))^2)$.

The maximizations take as input q_{ik}^m, ξ_{ik}^m and Υ_{ik}^m. These are computed conditional on each individuals' data and the m^{th} iteration parameter estimates.

$$\begin{aligned}
\xi_{ik}^m =& ((\Delta^m)^{-1} + D_i(\Sigma^m)^{-1})^{-1}((\Delta^m)^{-1}\gamma_k^m + D_i(\Sigma^m)^{-1}\overline{z_i}) \\
\Upsilon_{ik}^m =& ((\Delta^m)^{-1} + D_i(\Sigma^m)^{-1})^{-1} \\
q_{ik}^m =& \frac{\pi_k(g_i,\phi^m)L_c(d_i|\widetilde{X}_{it}(\theta^m,\gamma_k^m),\omega^m)\int_\mu L_w(w_i|\mu,\theta^m,\sigma^m)f(\mu|\gamma_k^m,\Delta^m)d\mu}{\sum_{k'=1}^{K}\pi_{k'}(g_i,\phi^m)L_c(d_i|\widetilde{X}_{it}(\theta^m,\gamma_{k'}^m),\omega^m)\int_\mu L_w(w_i|\mu,\theta^m,\sigma^m)f(\mu|\gamma_{k'}^m,\Delta^m)d\mu}
\end{aligned}$$

Where

$$\Sigma(j,j') = \left\{ \begin{array}{ll} \sigma_j^2 & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{array} \right\}$$

$$D_i(j,j') = \left\{ \begin{array}{ll} \sum_{t=1}^{T_i}(d_{it}=j) & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{array} \right\}$$

$$\overline{z}_i(j) = \left\{ \frac{\sum_{t=1}^{T_i}(d_{it}=j)(w_{it}-x_{it}\theta_j^m)}{\sum_{t=1}^{T_i}(d_{it}=j)} \right\}$$

C NLSY97 Activity Assignment and Sample Selection

Assignment of activities was done sequentially beginning with education.

Schooling The schooling assignment was based on reported attendance and did not require them to actually complete the grade. This is consistent with a general model where individual's make education choices with uncertain outcomes. Since the quasi-structural model only estimates policy functions conditional on state variables, we are not required to make assumptions on the how agents form expectations over outcomes. For primary and secondary education, the individual was coded as attending school if they reported attending any grade at all.

The NLSY97 provides a monthly postsecondary education enrollment indicator. In general it does not appear that this variable is informative about actual attendance as the mode number of months of attendance for those with positive attendance is 12 months. For those pursuing postsecondary education below an advanced degree, a monthly term variable was collected. This variable appears much more reliable as most individuals report no term over the summer months. Therefore, these individuals where assigned to schooling if they reported attending college and provided a valid term for at least 6 out of 12 months. Schooling was not assigned to those that did not have a valid term. However, for these individuals, if it was apparent that they were not enlisted in the military and not employed fulltime, then they were recoded as attending school.

Term information was not collected for those attending graduate school, so the monthly attendance variable was the sole criteria. If information is not available for the entire year, for example in their last interview round, then the information in available months was converted to a 12 month equivalent and the same criteria applied.

Employment If the respondent did not meet the school enrollment criteria and reported working at least 1,400 labor hours for the year, then the individual was assigned to employment. Again, if the individual was not present for the entire year, the weeks observed where converted to an annual equivalent. If the individual worked multiple full-time jobs, then the attributes of the job with the most fulltime weeks was assigned for that year. For each job, the respondent identifies the 3-digit 2002 Census occupation code. Given the reported 3-digit occupation, the individual is assigned to one of the 13 occupation categories following table 1.[12] Finally, wages are

[12]Previous papers modeling career decisions, (e.g. Neal (1999); Pavan (2009); Kambourov and Manovskii (2009)) using the NLSY79 or Panel Study of Income Dynamics (PSID) data have documented the high potential for measurement error when the reported occupation code is taken directly from the data. To avoid counting false career changes these papers impose a number of edits on the occupational data. The primary edit is to not consider any occupational change unless it is accompanied by a change in employer. Yamaguchi (2010) points out that this may be an undesirable restriction on

assigned using the hourly compensation rate of pay, which includes all forms of monetary compensation affiliated with the job and is deflated using the consumer price index to year 2008 dollars.

Military If an individual reported active military service for 25 or more weeks in a year their activity was coded as military.

Unemployment Finally, if they did not meet any of the assignment criteria, they were assigned to unemployed.

The analysis focuses on the white male cross-sectional sample from the data. This group consists of 2,284 individuals. The discrete decision period corresponds to the school year (September to August). The decisions of individuals are continually tracked from age 16 (t=1) until round 12. Of the original population, 2,138 remain in the survey until at least age 16. If the gap between interview dates for any individual exceeds 16 months, then the remaining observations for that individual are dropped from the analysis.[13] This leaves 17,616 person-years for the analysis.

Finally, an individual's remaining observations where dropped if any one of three events occurred: 1) their highest grade completed or degree completed is invalid from the survey data, 2) the individual's primary activity was employment, but no occupation was reported or the reported wage was greater than \$100 per hour or less than \$5 per hour[14], or 3) reported com-

---

the data as it is likely to exclude important career changes as individuals are promoted within the firm. The NLYS97 is unique from the NLSY79 and PSID in that it likely does not suffer from systematic measurement error in the reported occupation. From the beginning, the NLSY97 was conducted with a computer-assisted interview system which allows interviewers to reference back to the responses of their previous years interview. Interviewees are first read their previous years job description and are asked if that continues to define their job function. The occupation code only changes if they report a change in duties. Pavan (2009), using the NLSY79 data, cites evidence of spurious reported occupation changes by the fact that 40% of individuals remaining with the same employer in consecutive periods report a change in 3-digit occupations. The analogous figure for the NLSY97 data at the 3-digit detail level is only 13% of workers. This is a reasonable figure representing mobility within firms. The fact that this number is not overstated provides reasonable assurance of the reliability of the observed occupation changes. This feature makes the NLSY97 particularly desirable for a model that looks at a finer level of occupational choice.

[13]The exception is for interviews conducted in 1997. In many cases the gap between the round 1 and round 2 interview exceeded 18 months.

[14]In 2008 real dollars.

pleting 20 or more years of education.[15] These filters affected 7% of the records, leaving 2,117 individuals and 16,371 person-years for the analysis, averaging 7.66 years per individual.

D Parameter Estimates

Table 10: Logit Initial Condition Parameters $\pi(g_i|\phi)$

	Type 1	Type 2
Constant ϕ_{1k}	0.000	-6.738^{***}
	–	(1.236)
Mother HGC ϕ_{2k}	0.000	0.478^{***}
	–	(0.089)
Mother HGC Missing HGC ϕ_{3k}	0.000	6.292^{***}
	–	(1.260)
HGC at Age 16 LT 10th ϕ_{4k}	0.000	-0.653^{***}
	–	(0.258)

***,**,* Significantly different from zero at 1%, 5%, and 10% level respectively. Constructed from 800 non-parametric bootstraps of the model.

The 1,350 variables describing the policy functions are available from the author by request.

[15]There appear to be a large fraction of medical students who meet the employment criteria with 20 years of education. Mean wages increase monotonically across education levels except for 19 years to 20 years where the average wage falls 28%. These outliers have a strong effect on the returns to education, so these 20 observations are dropped.

Table 11: Wage Estimates

	Mgr	Bus	Prof 1	Prof 2	Serv 1	Food	Serv 2	Sales	Off-ice	Cons-truc	Main-ten	Prod	Tran
Educ. θ_j^s	0.053^{***} (0.011)	0.018 (0.017)	0.052^{***} (0.011)	0.096^{***} (0.014)	0.074^{***} (0.007)	0.030^{***} (0.011)	0.007 (0.007)	0.043^{***} (0.012)	0.022^{*} (0.011)	0.038^{***} (0.011)	0.023^{**} (0.011)	0.036^{***} (0.014)	0.023^{***} (0.009)
x_{it} MGR	0.104^{***} (0.022)	−0.002 (0.141)	0.089 (0.159)	0.088 (0.088)	0.896 (0.555)	−0.045 (0.104)	0.188^{**} (0.086)	0.111 (0.186)	0.106 (0.183)	0.092 (0.111)	0.243 (0.165)	−0.004 (0.084)	−0.101 (0.081)
x_{it} BUS	0.368^{*} (0.188)	0.080^{***} (0.025)	0.147^{*} (0.081)	0.000 –	0.406^{*} (0.243)	0.477^{**} (0.220)	0.557^{**} (0.270)	0.161 (0.104)	0.281 (0.184)	0.000 –	0.192 (0.119)	0.000 –	0.330 (0.249)
x_{it} PROF1	0.122^{***} (0.032)	0.157^{*} (0.086)	0.102^{***} (0.015)	0.278^{***} (0.103)	0.000 –	0.026 (0.204)	0.000 –	0.113 (0.170)	0.318^{***} (0.077)	0.076 (0.097)	0.205^{***} (0.061)	−0.024 (0.120)	0.000 –
x_{it} PROF2	0.062 (0.097)	0.038 (0.167)	0.068 (0.051)	0.027 (0.022)	0.005 (0.171)	0.243 (0.219)	−0.017 (0.123)	0.086 (0.115)	0.072 (0.157)	0.117 (0.161)	0.183 (0.194)	-0.181^{**} (0.089)	0.051 (0.198)
x_{it} SERV1	0.171^{**} (0.074)	0.329^{***} (0.119)	−0.078 (0.113)	0.153^{*} (0.081)	0.056^{***} (0.017)	0.274^{***} (0.108)	−0.071 (0.202)	0.107 (0.096)	0.095^{***} (0.039)	−0.040 (0.105)	0.013 (0.046)	−0.156 (0.237)	0.105 (0.139)
x_{it} FOOD	−0.020 (0.042)	−0.231 (0.155)	−0.169 (0.110)	−0.147 (0.144)	−0.081 (0.159)	0.050^{***} (0.014)	−0.035 (0.073)	−0.014 (0.037)	0.031 (0.038)	−0.013 (0.038)	0.034 (0.048)	0.067^{**} (0.033)	0.075^{**} (0.038)
x_{it} SERV2	0.312 (0.199)	-0.275^{***} (0.110)	−0.266 (0.181)	−0.021 (0.154)	0.018 (0.069)	0.040 (0.057)	0.061^{***} (0.020)	0.206^{**} (0.091)	−0.078 (0.056)	0.075 (0.047)	0.002 (0.058)	0.050 (0.048)	0.063 (0.097)
x_{it} SALES	0.077 (0.077)	0.069 (0.058)	−0.092 (0.061)	0.038 (0.104)	0.201 (0.242)	0.084 (0.062)	0.078 (0.048)	0.055^{***} (0.014)	0.038 (0.027)	−0.049 (0.032)	0.027 (0.058)	0.046 (0.041)	0.053^{*} (0.032)
x_{it} OFFICE	0.038 (0.038)	0.003 (0.060)	0.037 (0.082)	0.060 (0.118)	0.006 (0.087)	0.042 (0.089)	0.023 (0.077)	0.066^{***} (0.020)	0.061^{***} (0.012)	0.015 (0.061)	−0.020 (0.049)	0.022 (0.044)	0.042^{**} (0.019)
x_{it} CONST	0.081 (0.053)	−0.128 (0.544)	0.144 (0.096)	−0.029 (0.138)	0.026 (0.040)	0.029 (0.064)	0.160^{***} (0.057)	0.082 (0.058)	0.042 (0.031)	0.075^{***} (0.008)	0.042 (0.049)	0.012 (0.025)	0.091^{***} (0.037)
x_{it} MAINT	0.264^{*} (0.149)	0.030 (2.790)	0.070 (0.065)	0.079 (0.115)	−0.056 (0.163)	0.179 (0.163)	0.151 (0.156)	0.116 (0.076)	0.235^{***} (0.042)	0.041 (0.027)	0.083^{***} (0.017)	0.063 (0.041)	0.085^{**} (0.042)
x_{it} PROD	−0.016 (0.091)	−0.022 (0.173)	−0.004 (0.109)	−0.127 (0.133)	−0.018 (0.078)	0.053 (0.045)	0.015 (0.064)	0.048 (0.075)	0.042 (0.029)	−0.006 (0.038)	0.025 (0.024)	0.071^{***} (0.009)	0.028^{*} (0.015)
x_{it} TRAN	0.066 (0.042)	0.119 (0.231)	−0.063 (0.094)	0.173^{*} (0.099)	0.003 (0.065)	0.062 (0.044)	0.091 (0.062)	0.092^{***} (0.030)	0.048^{*} (0.026)	0.047 (0.029)	0.067^{**} (0.030)	0.029 (0.021)	0.054^{***} (0.009)
σ_j^2	0.080^{***} (0.017)	0.054^{***} (0.012)	0.053^{***} (0.013)	0.046^{***} (0.010)	0.029^{***} (0.006)	0.041^{***} (0.011)	0.023^{***} (0.006)	0.065^{***} (0.016)	0.037^{***} (0.008)	0.067^{***} (0.010)	0.072^{***} (0.015)	0.035^{***} (0.005)	0.047^{***} (0.008)

***,**,* Significantly different from zero at 1%, 5%, and 10% level respectively, constructed from 800 non-parametric bootstraps of the model.
$^{-}$ Not Identified. No workers with this experience in this occupation.